

SAMPLING ALGEBRAIC VARIETIES FOR SUM OF SQUARES PROGRAMS

DIEGO CIFUENTES AND PABLO A. PARRILO

ABSTRACT. We study sum of squares (SOS) relaxations to decide the nonnegativity of a polynomial p on the set $\mathcal{V} \cap \mathbb{R}^n$, where \mathcal{V} is a complex algebraic variety. We propose a new methodology that, rather than relying on some algebraic description, represents \mathcal{V} with a generic set of complex samples. This approach depends only on the geometry of \mathcal{V} , avoiding representation issues such as multiplicity and choice of generators. It also takes advantage of the dependencies in the coordinate ring of \mathcal{V} to reduce the size of the corresponding semidefinite program (SDP). In addition, the input can be given as a straight-line program. Our methods are particularly appealing for varieties which are easy to sample from but for which the defining equations are complicated, such as $SO(n)$, Grassmannians or rank k tensors. Nonetheless, for arbitrary varieties we can obtain the samples by using the tools of numerical algebraic geometry. In this way we connect the areas of SOS optimization and numerical algebraic geometry.

1. INTRODUCTION

Consider the ring $\mathbb{R}[x] := \mathbb{R}[x_1, \dots, x_n]$ of multivariate polynomials. For some polynomial $p \in \mathbb{R}[x]$, we are interested in deciding whether

$$(1) \quad p(x) \geq 0 \text{ for all } x \in \mathcal{V} \cap \mathbb{R}^n,$$

where \mathcal{V} is a complex algebraic variety. Recall that an algebraic variety is the zero set of some set of polynomials $h = \{h_1, \dots, h_m\} \subseteq \mathbb{R}[x]$, i.e.,

$$\mathcal{V} = \mathcal{V}_{\mathbb{C}}(h) = \{x \in \mathbb{C}^n : h_1(x) = \dots = h_m(x) = 0\}.$$

The above decision problem is computationally hard, but there are simpler relaxations based on *sum of squares* [16, 18]. A polynomial $F \in \mathbb{R}[x]$ is a sum of squares (SOS) if it can be written in the form $F(x) = \sum_i f_i(x)^2$ for some $f_i \in \mathbb{R}[x]$. An SOS certificate of the nonnegativity of $p(x)$ on $\mathcal{V} \cap \mathbb{R}^n$ is a tuple of polynomials (F, g_1, \dots, g_m) , where $F(x)$ is SOS and

$$(2) \quad p(x) = F(x) + \sum_j g_j(x)h_j(x).$$

Given a bound d on the degrees of these polynomials, finding an SOS certificate reduces to a semidefinite program (SDP).

Assuming that an SOS certificate exists, the minimal degree bound needed naturally depends on the polynomial p and the variety \mathcal{V} . However, it also strongly depends on the specific description of the variety, i.e., the set of equations h we use. Indeed, for

Date: November 23, 2015.

Key words and phrases. Sum of squares, Sampling, SDP, Numerical Algebraic Geometry.

a given polynomial $f(x) = p(x) - F(x)$ the minimal degrees of g_j such that $f(x) = \sum_j g_j(x)h_j(x)$ are strongly dependent on the generators h_j used. The multiplicity structure of h also plays an important role in SOS certificates, possibly increasing the complexity. Furthermore, if we do not have explicit equations describing \mathcal{V} , or if we do not want to expand these equations, the above relaxation cannot be easily applied.

In this paper we propose an alternative geometric approach to certify nonnegativity on $\mathcal{V} \cap \mathbb{R}^n$. Concretely, rather than depending on an algebraic description, we rely on a generic set of samples on \mathcal{V} . Naturally, this approach is likely to be more efficient when there is a simple procedure to sample points from the variety. Note that such procedure exists for many interesting varieties, such as the set of orthogonal matrices, $SO(n)$, Grassmannians, rank k tensors and more general secant varieties. We remark that secant varieties are a rich class of problems for which computing the defining equations is typically prohibitive [12]. Moreover, for an arbitrary variety \mathcal{V} the field of *numerical algebraic geometry* provides practical methods to sample generic points [23]. This is a recent subfield in computational algebraic geometry that offers several numerical and computational advantages over traditional symbolic methods such as Gröbner bases.

Sampling certificates. We now present the main idea of this paper. Let $Z := \{z_1, \dots, z_S\} \subseteq \mathcal{V}$ be a set of points on the variety. Evaluating equation (2) in these samples we obtain:

$$(3) \quad p(z_s) = F(z_s), \quad \text{for } s = 1, \dots, S.$$

As before, computing an SOS polynomial F satisfying the above equations reduces to an SDP. We refer to a pair (F, Z) satisfying (3) as a *sampling SOS certificate* (see Definition 3.1). We show that under certain genericity assumptions equations (2) and (3) are equivalent and thus, a sampling SOS certificate is a valid certificate of nonnegativity. Note, however, that equation (3) is much simpler as it does not contain the polynomials h_j defining the variety. In fact, these polynomials are not needed as we are effectively representing the variety \mathcal{V} with the set of samples. As the generators h_j are now implicit, our methodology does not depend on the algebraic properties of these equations, but only on the geometry of the set.

Our methods extend the ideas from Löfberg and Parrilo in [15], where they first consider sampling formulations for unconstrained SOS problems. They show that sampling formulations offer some numerical advantages over the standard approach. Most remarkably, the SDP has a low rank structure, which leads to a significant complexity improvement in interior point methods. In particular, low rank structure is exploited in the solvers SDPT3 and DSDP [3, 25]. Secondly, the SDP is usually better conditioned, as it relies on a set of orthogonal polynomials instead of a monomial basis. These properties make sampling formulations appealing, as seen in [14, 20, 21]. We show that these properties are preserved in the variety case considered in this paper. We remark that our use of samples differs from [15] in that they not only lead to an alternative SDP formulation, but they also describe the underlying variety \mathcal{V} .

Our methodology also shares some similarities with *quotient ring* SOS formulations based on Gröbner bases, introduced in [17]. As in equation (3), the quotient ring SDP does not contain the polynomials h_j , as these polynomials are implicit in the SDP. Moreover, this formulation takes advantage of the algebraic dependencies derived from

the equations h_j to reduce the size of the SDP. A drawback of this approach is that we need a description of the quotient ring $\mathbb{R}[x]/I$, where $I = \langle h \rangle$, and the Gröbner bases approach is typically too expensive. We will see that such description can also be obtained by using a suitable set of samples of the variety. Therefore, our approach can be seen as a quotient ring formulation that is based on samples, instead of Gröbner bases. We remark that unlike standard quotient ring formulations we ignore multiplicity structure, and thus we are actually working in the coordinate ring of the variety.

Contributions and outline. This paper presents the following contributions.

- We introduce a new methodology to SOS over an algebraic variety \mathcal{V} . This is a geometric formulation that represents \mathcal{V} with a generic set of complex samples, instead of relying on some algebraic description. In this way we avoid algebraic issues such as multiplicity and the dependence on the specific generators used. We analyze the correctness of our formulation, establishing sufficient conditions on the samples and the variety.
- We show that our methodology can be seen as a quotient ring SOS formulation that is based on samples, in contrast to the standard Gröbner bases approach. In particular, our formulation takes advantage of the algebraic dependencies of the coordinate ring to reduce the size of the SDP. This makes our methods appealing for many varieties which are easy to sample from but for which Gröbner bases computation is intractable. Examples of such varieties include $SO(n)$, Stiefel manifolds, Grassmannians and secant varieties.
- We apply for the first time techniques from numerical algebraic geometry to SOS programs. In this way, we inherit some of the main strengths from this area. In particular, we highlight the intrinsic parallelism of these methods. Furthermore, this allows us to work with straight-line programs, i.e., we do not need to expand the polynomials.

The structure of this paper is as follows. Section 2 presents some basic algebraic preliminaries. Section 3 introduces the main object of this paper, sampling SOS certificates, and we provide an SDP that computes them, given a set of samples. Section 4 analyzes the conditions required on the samples, and the methods we propose to select them. Section 5 presents several examples to illustrate our methods. Appendix A briefly describes two traditional SOS methodologies, comparing them to our approach.

2. PRELIMINARIES

2.1. Algebraic geometry. Let \mathbb{K} denote a field which is either \mathbb{R} or \mathbb{C} , and let $\mathbb{K}[x] = \mathbb{K}[x_1, \dots, x_n]$ denote the ring of polynomials with coefficients in \mathbb{K} . The *ideal* generated by a set of polynomials $h = \{h_1, \dots, h_m\} \subseteq \mathbb{K}[x]$ is

$$I = \langle h \rangle := \{ \sum_i g_i h_i : g_i \in \mathbb{K}[x] \}.$$

Given an ideal I , its *quotient ring* $\mathbb{K}[x]/I$ is the set of equivalence classes where $f \sim g$ if $f - g \in I$.

Given a set of polynomials $h \subseteq \mathbb{K}[x]$, its complex *algebraic variety* is

$$\mathcal{V} = \mathcal{V}_{\mathbb{C}}(h) := \{x \in \mathbb{C}^n : h_i(x) = 0 \text{ for } h_i \in h\}.$$

The corresponding real variety is $\mathcal{V} \cap \mathbb{R}^n$. Observe that $\mathcal{V}_{\mathbb{C}}(h) = \mathcal{V}_{\mathbb{C}}(\langle h \rangle)$. In this paper we only consider complex varieties defined by real polynomials. It is easy to see that a variety $\mathcal{V} \subseteq \mathbb{C}^n$ is defined by real polynomials if and only if it is *self-conjugate*, i.e. its complex conjugate $\overline{\mathcal{V}}$ is itself.

The *coordinate ring* of a variety $\mathcal{V} \subseteq \mathbb{C}^n$ is the quotient ring $\mathbb{K}[\mathcal{V}] := \mathbb{K}[x]/I_{\mathbb{K}}(\mathcal{V})$, where $I_{\mathbb{K}}(\mathcal{V})$ is the vanishing ideal

$$I_{\mathbb{K}}(\mathcal{V}) := \{f \in \mathbb{K}[x] : f(x) = 0 \text{ for all } x \in \mathcal{V}\}.$$

Equivalently, $\mathbb{K}[\mathcal{V}]$ is the set of equivalence classes of polynomials where $f \sim g$ if they define the same function on \mathcal{V} .

Remark. Let \mathcal{V} be the variety of an ideal I . Recall that Hilbert Nullstellensatz says that $I_{\mathbb{K}}(\mathcal{V}) = \sqrt{I}$. Therefore, $\mathbb{K}[\mathcal{V}]$ is equal to the quotient ring $\mathbb{K}[x]/I$ only if I is radical.

We say that a variety $\mathcal{V} \subseteq \mathbb{C}^n$ is *irreducible* if it is not the union of two proper algebraic varieties. The closure of the image of an irreducible variety under a rational map is also irreducible. In particular, any variety parametrized by \mathbb{C}^n is irreducible.

A variety $\mathcal{V} \subseteq \mathbb{C}^n$ can be decomposed in a unique way in the form

$$(4) \quad \mathcal{V} = \mathcal{V}_1 \cup \dots \cup \mathcal{V}_r$$

where each \mathcal{V}_i is an irreducible variety which is not contained in any other \mathcal{V}_j . The varieties \mathcal{V}_i are called the *irreducible components* of \mathcal{V} . If \mathcal{V} is self-conjugate, then either \mathcal{V}_i is also self-conjugate, or there is a pair $(\mathcal{V}_i, \mathcal{V}_j)$ of conjugate components.

2.2. Genericity, randomness and identity testing. The notion of *genericity* is fundamental in algebraic geometry. Let $\mathcal{V} \subseteq \mathbb{C}^n$ be an irreducible variety with infinitely many points and let z denote a sample point on \mathcal{V} . We say that z satisfies a property *generically* if there is a nonzero polynomial $q \in \mathbb{C}[\mathcal{V}]$ such that the property holds except perhaps when $q(z) = 0$. Informally, this means that the property holds outside of the small bad region given by $q(z) = 0$. We will slightly modify this notion of genericity in Section 4.1.

We often say that $z \in \mathcal{V}$ is *generic* if it satisfies generically some property of interest (such as the conclusion of a theorem). A generic point can be understood as a random point on the variety. For instance, consider the following statement.

Proposition 1. *Let \mathcal{V} be an irreducible variety, let $f \in \mathbb{C}[\mathcal{V}]$ be a nonzero polynomial and let $z \in \mathcal{V}$ be a generic sample. Then $f(z)$ is nonzero.*

Proof. The conclusion holds except in the bad region defined by $f(z) = 0$. \square

Genericity allows us to derive *randomized* algorithms that succeed with *probability one* with respect to any distribution on \mathcal{V} with full support. A prototypical example is the *identity testing* problem, of determining whether some polynomial $f(x)$ is identically zero on \mathcal{V} . Note that f can be given as a straight-line program (such as a determinant), so the problem is nontrivial even when $\mathcal{V} = \mathbb{C}^n$. Proposition 1 gives rise to Algorithm 1. This method efficiently solves the identity testing problem, provided that we can sample the variety. Surprisingly, no efficient deterministic algorithm to this problem is known, and it is likely that finding such algorithm is very hard [11].

Algorithm 1 Identity testing over \mathbb{C}

Input: Polynomial $f \in \mathbb{C}[x]$, irreducible variety \mathcal{V} **Output:** “True”, if f is identically zero on \mathcal{V} . “False”, otherwise.

```

1: procedure IsZERO( $f, \mathcal{V}$ )
2:    $z :=$  generic sample from  $\mathcal{V}$ 
3:   if  $f(z) = 0$  then
4:     return True
5:   return False

```

The sampling SOS certificates described in this paper depend on the choice of some generic samples, and thus are randomized in nature. This contrasts with standard SOS certificates which are deterministic. We will also see that Algorithm 1 can be used to verify the validity of a sampling SOS certificate.

Remark. Although these randomized algorithms provably work with probability one in exact arithmetic, further numerical considerations have to be taken into account when working in floating point; see e.g. [23, §4].

2.3. Sampling varieties. Our technique requires a procedure to sample points of a complex variety \mathcal{V} . More precisely, we need to sample generic points in each irreducible component of \mathcal{V} . We note that if \mathcal{V} is the image of some variety \mathcal{W} under some map ψ , then we just need to sample points on \mathcal{W} . In particular, if \mathcal{V} is parametrized by \mathbb{C}^n then the sampling problem is trivial. The difficult case is when \mathcal{V} is described as the zero set of some polynomials $h \subseteq \mathbb{C}[x]$. A practical approach to accomplish this is through the tools of numerical algebraic geometry.

Let \mathcal{V} be decomposed as in (4). A *numerical irreducible decomposition* of \mathcal{V} is a set $\{W_1, \dots, W_r\}$, where W_i is a *witness set* of the irreducible component \mathcal{V}_i . A witness set W_i is a numerical data structure that represents component \mathcal{V}_i , which can be used to do several several computations on it. In particular, given W_i we can sample an arbitrary number of generic points in \mathcal{V}_i . This functionality is available in homotopy continuation tools such as Bertini [1] and PHCpack [26].

We finally highlight some of the advantages of numerical methods with respect to standard symbolic methods, such as Gröbner bases:

- they are naturally parallelizable.
- they allow for straight-line programs, i.e. polynomials do not need to be expanded.
- they offer better numerical stability.

We refer the reader to [2, 23] for more information about numerical algebraic geometry.

2.4. Polynomial interpolation. Let $\mathcal{V} \subseteq \mathbb{C}^n$ be a self-conjugate variety and let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$ be its coordinate ring. Let $\mathcal{L} \subseteq \mathcal{R}$ be a finite dimensional subspace of polynomials. Our methodology represents a variety \mathcal{V} in terms of a set of samples $Z \subseteq \mathcal{V}$. In order for this approach to work we need the samples to extrapolate well the variety. More precisely, we want the samples Z to uniquely determine the polynomials of \mathcal{L} . The formal statement is presented now.

Definition 2.1. Let $\mathcal{V} \subseteq \mathbb{C}^n$ be a self-conjugate variety and let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$. Let $\mathcal{L} \subseteq \mathcal{R}$ be a linear subspace and let $Z \subseteq \mathcal{V}$ be a set of samples. We say that (\mathcal{L}, Z) is *poised*¹ if the only polynomial $q \in \mathcal{L}$ such that $q(z) = 0$ for all $z \in Z$ is the zero polynomial.

Once we fix a set of generators of \mathcal{L} , we can equivalently state the poisedness condition in terms of matrices, as we explain now. Let $u(x) \in \mathcal{R}^N$ be a polynomial vector whose entries span \mathcal{L} . Let U be the matrix with columns $u(z)$ for $z \in Z$ and let $\hat{U} := [U | \overline{U}]$. We will refer to the (complex) rank of matrix \hat{U} as the *empirical dimension* of \mathcal{L} with respect to Z . It is easy to see that the empirical dimension does not depend on the specific set of generators used.

Lemma 2. *Let $\mathcal{L} \subseteq \mathcal{R}$ and $Z \subseteq \mathcal{V}$ be as above. Then (\mathcal{L}, Z) is poised if and only if the dimension of \mathcal{L} is equal to its empirical dimension.*

Proof. Let D be the dimension of \mathcal{L} , and let $u(x) \in \mathcal{R}^D$ be a basis. Let U be the matrix with columns $u(z)$ for $z \in Z$, and let $\hat{U} = [U | \overline{U}]$. Let's assume that the rank of \hat{U} is equal the true dimension D . Note that any $q \in \mathcal{L}$ can be written uniquely in the form $q(x) = \mu^T u(x)$ for some $\mu \in \mathbb{R}^D$. The condition that $q(z) = 0$ for $z \in Z \cup \overline{Z}$ implies that $\mu^T \hat{U} = 0$. As \hat{U} has full row rank then $\mu = 0$, and thus (\mathcal{L}, Z) is poised.

Assume now that \hat{U} does not have full row rank. Then there is some nonzero $\lambda \in \mathbb{C}^D$ such that $\lambda^T \hat{U} = 0$. Observe that this implies that $\Re(\lambda)^T U = \Im(\lambda)^T U = 0$. Thus, there is a nonzero $\mu \in \mathbb{R}^D$ such that $\mu^T U = 0$. By considering the polynomial $q(x) := \mu^T u(x)$ we conclude that (\mathcal{L}, Z) is not poised. \square

Example 2.1. Let $\mathcal{V} = \mathbb{C}$ and $\mathcal{R} = \mathbb{R}[x]$ be the space of univariate polynomials. Let \mathcal{L} be the set of polynomials of degree less than N and let $u(x) = (1, x, \dots, x^{N-1})$. Let $Z \in \mathbb{C}^{N/2}$ be a tuple of complex samples, and let \hat{Z} be the concatenation of Z and \overline{Z} . The matrix \hat{U} in this case is the Vandermonde matrix of \hat{Z} . We know that this matrix is singular only if there are repeated elements in \hat{Z} . Therefore, (\mathcal{L}, Z) is poised if and only if the elements of \hat{Z} are all distinct.

3. SAMPLING SOS CERTIFICATES

In this section we introduce our main object of study, S-SOS certificates (sampling SOS certificates), and we show a semidefinite program that computes them.

Definition 3.1. Let $\mathcal{V} \subseteq \mathbb{C}^n$ be a self-conjugate variety and let $p \in \mathbb{R}[x]$ be nonnegative on $\mathcal{V} \cap \mathbb{R}^n$. An *S-SOS pre-certificate* is a pair (F, Z) where $F \in \mathbb{R}[x]$ is a sum of squares (SOS) and $Z = \{z_1, \dots, z_S\} \subseteq \mathcal{V}$ is a set of samples such that

$$(5) \quad p(z_s) = F(z_s) \quad \text{for } s = 1, \dots, S.$$

We say that the pre-certificate is *correct* if furthermore

$$(6) \quad p(z) = F(z) \quad \text{for all } z \in \mathcal{V}.$$

An *S-SOS certificate* is a correct pre-certificate.

¹In polynomial interpolation it is usually further required that $|Z| = D$, where D is the dimension of \mathcal{L} [22]. We do not impose such condition.

Remark. Note that an S-SOS certificate (F, Z) is a valid certificate of nonnegativity, as $p(x) = F(x) \geq 0$ for all $x \in \mathcal{V} \cap \mathbb{R}^n$.

Computing S-SOS certificates is the main problem we address in this paper. We will divide this problem into three steps.

- (i) Obtain a good set of samples Z on the variety. We will see that choosing generic samples is enough.
- (ii) Given Z , compute an S-SOS pre-certificate (F, Z) using an SDP.
- (iii) Verify that the pre-certificate (F, Z) is correct.

This section is concerned mainly with the second problem: computing pre-certificates. We will also see what do we mean by a “good” set of samples, and the answer has to do with the poisedness property from Section 2.4. The first and third problems will be discussed in Section 4.

3.1. Sampling SDP. We now describe the procedure to obtain a pre-certificate given the sample set Z , i.e., how to compute an SOS polynomial F satisfying equation (5). We will see that this problem reduces to a semidefinite program (SDP) in a similar manner of standard SOS methods.

Recall that a polynomial $F \in \mathbb{R}[x]$ is SOS if and only if

$$F(x) = Q \bullet u(x)u(x)^T$$

for some vector of monomials $u(x) \in \mathbb{R}[x]^N$ and some positive semidefinite matrix Q (denoted $Q \succeq 0$), where the notation \bullet is for trace inner product [16, 18]. Typically $u(x)$ consists of all $N = \binom{n+d}{d}$ monomials of degree at most d , where d is some given degree bound. Then computing an SOS polynomial F satisfying (5) reduces to the following SDP:

$$(7) \quad \boxed{\begin{array}{ll} \text{find} & Q \in \mathbb{R}^{N \times N}, \quad Q \succeq 0 \\ \text{subject to} & p(z_s) = Q \bullet u(z_s)u(z_s)^T, \quad \text{for } s = 1, \dots, S \end{array}}$$

Note that the matrix Q is real, whereas $p(z_s)$ and $u(z_s)$ are complex. Thus, each equality imposes a constraint on both the real and the imaginary part, i.e.,

$$\Re(p(z_s)) = Q \bullet \Re(u(z_s)u(z_s)^T), \quad \Im(p(z_s)) = Q \bullet \Im(u(z_s)u(z_s)^T).$$

The above SDP has two important features: the polynomial p can be given as a *straight-line program* (i.e., it does not need to be expanded) and the constraint matrices have *low rank*. Indeed, the rank of the constraint matrices $\Re(u(z_s)u(z_s)^T)$ and $\Im(u(z_s)u(z_s)^T)$ is at most two. This special rank structure can be exploited in interior point methods, as discussed in [3, 15, 21]. In particular, the Hessian assembly takes only $O(N^3)$ for low rank matrices, as opposed to $O(N^4)$ for unstructured matrices.

Observe that the monomial vector $u(x)$ can be replaced by any other polynomial set with the same linear span. In particular, we will see in Section 3.3 that $u(x)$ can be chosen to be an orthogonal basis with respect to a natural inner product supported on the samples. Remarkably, this orthogonalization reduces complexity in the SDP by exploiting the algebraic *dependencies* of the coordinate ring $\mathbb{R}[\mathcal{V}]$. In addition, the

conditioning of the problem might improve, as explained in [15] for the unconstrained case $\mathcal{V} = \mathbb{C}^n$.

Remark (Kernel/Image form). The feasible set of (7) has the form $Q \succeq 0, Q \in \mathcal{Q}$, where

$$\mathcal{Q} = \{Q : Q \bullet A_i = b_i\}$$

is an affine subspace. We refer to the above representation of \mathcal{Q} as the *kernel form*. Alternatively, we can describe \mathcal{Q} explicitly by giving some generators, i.e.,

$$\mathcal{Q} = \{Q_0 + \sum_j \lambda_j Q_j : \lambda_j \in \mathbb{R}\}$$

where $Q_0 \bullet A_i = b_i$ and $Q_j \bullet A_i = 0$. We refer to this representation as the *image form*. Depending on the problem, either of such representations might be more convenient. In particular, if the number of constraints is close to the dimension of Q then the latter representation is more compact. This will be the case in the applications shown in Sections 5.2 and 5.3. For a given problem, we can decide which representation is better by estimating the number of variables used in both of them, as discussed in [18].

3.2. Poisedness implies correctness. We just showed how to compute an S-SOS pre-certificate for a given sample set. However, this pre-certificate might not be correct unless we are cautious with the sample set, as illustrated in the next example.

Example 3.1 (Incorrect pre-certificate). Let $\mathcal{V} \subseteq \mathbb{C}^2$ be the zero set of $h(x) := x_2^2 - 1$ and let $p(x) := x_1^2 - x_2 + 1$, which is nonnegative on $\mathcal{V} \cap \mathbb{R}^2$. Let $Z := \{(i, 1)\}_{i=1}^S \subseteq \mathcal{V}$ be a set of samples and let $F(x) := x_1^2$. Observe that (F, Z) is an S-SOS pre-certificate, but it is not correct because $p(0, -1) \neq F(0, -1)$.

The reason why the above example failed is because the sample set did not capture well the geometry of the variety. We now present a sufficient condition that guarantees the correctness of an S-SOS pre-certificate. Let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$ be the coordinate ring of the variety, which is the space where we will work on. In particular, we will see the entries of the polynomial vector $u(x)$, as well as $p(x)$, as elements of \mathcal{R} .

We denote as $\mathcal{L}_1 \subseteq \mathcal{R}$ the linear space of polynomials spanned by the entries of $u(x)$. We also consider the linear space $\mathcal{L}_2 \subseteq \mathcal{R}$ spanned by the entries of $u(x)u(x)^T$. Note that the SOS polynomial $F(x) = Q \bullet u(x)u(x)^T$ lies in this space.

The following proposition tells us that the poisedness condition from Section 2.4 guarantees the correctness of an S-SOS pre-certificate. Therefore, given the linear subspace \mathcal{L}_2 , a *good set of samples* is such that (\mathcal{L}_2, Z) is *poised*. For the rest of this section we assume that this condition is satisfied. In Section 4 we will discuss how to choose the samples in order to satisfy this requirement.

Proposition 3. *Let $\mathcal{V} \subseteq \mathbb{C}^n$ be a self-conjugate variety, let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$ and let $p \in \mathcal{R}$ be nonnegative on $\mathcal{V} \cap \mathbb{R}^n$. Let (F, Z) be an S-SOS pre-certificate and let $\mathcal{L}_2 \subseteq \mathcal{R}$ be a linear subspace such that $p, F \in \mathcal{L}_2$. If (\mathcal{L}_2, Z) is poised then (F, Z) is correct.*

Proof. Let $g := p - F \in \mathcal{L}_2$, and observe that $g(z) = 0$ for $z \in Z$. As (\mathcal{L}_2, Z) is poised, this implies that $g = 0$ and thus $p = F \in \mathcal{R}$. \square

3.3. Basis selection. The polynomial vector $u(x)$ typically consists of all monomials up to some degree. These monomials are usually linearly dependent in the coordinate ring \mathcal{R} , so it is useful to obtain a basis of its linear span \mathcal{L}_1 before solving the SDP. We explain now how to get an orthogonal basis $u^o(x)$ with respect the inner product given in the next proposition. By using this basis we take into account the coordinate ring structure, which is why our sampling methodology is analogous to the quotient ring formulation based on Gröbner bases (see Appendix A.2 for further discussion).

Proposition 4. *Let $\mathcal{V} \subseteq \mathbb{C}^n$ be a self-conjugate variety and let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$. Let $\mathcal{L}_1 \subseteq \mathcal{R}$ be a linear subspace and let $Z \subseteq \mathcal{V}$ be a set of samples. Let $\langle \cdot, \cdot \rangle_Z : \mathcal{L}_1 \times \mathcal{L}_1 \rightarrow \mathbb{R}$ be*

$$\langle f, g \rangle_Z = \sum_{z \in Z} (f(z)g(\bar{z}) + f(\bar{z})g(z)).$$

If (\mathcal{L}_1, Z) is poised then $(\mathcal{L}_1, \langle \cdot, \cdot \rangle_Z)$ is a real inner product space.

Proof. It is clear that $\langle \cdot, \cdot \rangle_Z$ is bilinear and symmetric. Thus, we only need to check positiveness. Observe that $\langle f, f \rangle_Z = \sum_{z \in Z} 2|f(z)|^2 \geq 0$, which is zero only if $f(z) = 0$ for all $z \in Z$. As $f \in \mathcal{L}_1$, the poisedness condition implies $f = 0$. \square

Remark. Note that if (\mathcal{L}_2, Z) is poised then (\mathcal{L}_1, Z) is also poised.

To find the orthogonal basis, we will operate on the evaluation matrix U with columns $u(z)$ for $z \in Z$. Consider the real matrix $W := [\Re(U) | \Im(U)]$. Observe that $u(x)$ is an orthogonal basis with respect to $\langle \cdot, \cdot \rangle_Z$ if and only if the rows of W are orthogonal with respect to the standard real inner product. Thus, we just need to orthogonalize the rows of W . Using an SVD (or rank revealing QR), we can obtain a decomposition $W = TW^o$, where W^o has orthogonal rows and T is a real full rank transformation matrix. Let U^o be such that $W^o = [\Re(U^o) | \Im(U^o)]$. The matrix U^o encodes the new vector of orthogonal polynomials $u^o(x)$. We note that directly orthogonalizing the matrix U does not work, as the transformation matrix T would be complex.

Algorithm 2 Orthogonal basis on the coordinate ring

Input: Polynomial vector $u(x)$, samples Z of variety \mathcal{V}

Output: Orthogonal basis $u^o(x)$ and its evaluation matrix U^o .

- 1: **procedure** ORTHBASIS($u(x), Z$)
 - 2: $U :=$ evaluation matrix with columns $u(z)$ for $z \in Z$
 - 3: $W := [\Re(U) | \Im(U)]$
 - 4: orthogonalize $W =: TW^o$, where $W^o(W^o)^T = \text{id}$
 - 5: let U^o be such that $W^o = [\Re(U^o) | \Im(U^o)]$
 - 6: let $u^o(x)$ be such that $u(x) = Tu^o(x)$
 - 7: **return** $u^o(x), U^o$
-

Example 3.2. Let \mathcal{V} be the complex variety of the set of rotation matrices $SO(2)$, i.e.,

$$(8) \quad \mathcal{V} = \{X \in \mathbb{C}^{2 \times 2} : X^T X = \text{id}_2, \det(X) = 1\}.$$

Let $p(X) = 4X_{21} - 2X_{11}X_{22} - 2X_{12}X_{21} + 3$, which is nonnegative on $\mathcal{V} \cap \mathbb{R}^{2 \times 2}$. We want to find an S-SOS certificate. We can sample points in \mathcal{V} using the Cayley parametrization: $A \mapsto (I - A)(I + A)^{-1}$, for A skew-symmetric. Consider the following 3 complex samples:

$$z_1 = \begin{bmatrix} -0.6+0.8i & 1.2+0.4i \\ -1.2-0.4i & -0.6+0.8i \end{bmatrix}, \quad z_2 = \begin{bmatrix} -1.2+0.4i & 0.6+0.8i \\ -0.6-0.8i & -1.2+0.4i \end{bmatrix}, \quad z_3 = \begin{bmatrix} -0.75+0.25i & 0.75+0.25i \\ -0.75-0.25i & -0.75+0.25i \end{bmatrix}.$$

We fix a degree bound of $d = 1$, and we let $u(x) = (1, X_{11}, X_{12}, X_{21}, X_{22})$ be the monomials of degree at most d . The matrix of evaluations is:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ -0.6+0.8i & -1.2+0.4i & -0.75+0.25i \\ -1.2-0.4i & -0.6-0.8i & -0.75-0.25i \\ 1.2+0.4i & 0.6+0.8i & 0.75+0.25i \\ -0.6+0.8i & -1.2+0.4i & -0.75+0.25i \end{bmatrix}$$

Using an SVD we obtain the orthogonalized matrix U^o and the corresponding polynomial basis $u^o(x)$. Note that $u^o(x)$ has only 3 elements, as opposed to $u(x)$.

$$U^o = \begin{bmatrix} -0.5955+0.1005i & -0.5955-0.1005i & -0.5201 \\ 0.3058+0.6116i & -0.3058+0.6116i & 0.2548i \\ -0.0708+0.6411i & -0.0708-0.6411i & 0.4100 \end{bmatrix} \quad u^o(X) = (X_{21} + X_{22} - .8054, X_{21} - X_{22}, X_{21} + X_{22} + 2.4831)$$

The sampling SDP is

$$\begin{aligned} & \text{find } Q \in \mathbb{R}^{3 \times 3}, \quad Q \succeq 0 \\ & \text{subject to } p(z_s) = Q \bullet u_s u_s^T, \quad \text{for } s = 1, 2, 3 \end{aligned}$$

where u_s denotes the s -th column of U^o . Solving the SDP we obtain the S-SOS pre-certificate (F, Z) , where $F(X) = (2X_{21} + 1)^2$.

4. SELECTING THE SAMPLES

We now know how to compute an S-SOS pre-certificate given a sample set Z , and we also know that a good sample set must be such that (\mathcal{L}_2, Z) is poised. There are still two important questions that we need to address to produce an S-SOS certificate.

- (i) Given $\mathcal{L}_2 \subseteq \mathcal{R}$, how can we get a sample set Z such that (\mathcal{L}_2, Z) is poised?
- (iii) How do we verify that a pre-certificate (F, Z) is correct?

In order to answer both questions we will make use of randomness. More precisely, we will sample generic points on each irreducible component of \mathcal{V} . These samples can be obtained as explained in Section 2.3.

Let's show the answer to question (iii). In order for a pre-certificate (F, Z) to be correct, we need to check whether the polynomial $g := p - F$ is identically zero on \mathcal{V} . This is equivalent to showing that g is identically zero on each irreducible component $\mathcal{W} \subseteq \mathcal{V}$, and this can be done by using Algorithm 1. Therefore, there is a randomized algorithm that checks the correctness of an S-SOS certificate.

In this section we will show how a generic set of samples can also be used to satisfy the poisedness property. This will complete our methodology for producing (and verifying) an S-SOS certificate.

4.1. How many samples. As we decided that the samples will be random, the only missing point is to determine how many samples to take. Let D be the dimension of the linear subspace $\mathcal{L}_2 \subseteq \mathcal{R}$. Recall from Lemma 2 that (\mathcal{L}_2, Z) is poised if the empirical dimension, which is the rank of the evaluation matrix \hat{U}_2 , is equal to the

actual dimension D . Naturally, this implies that we need at least $\lceil D/2 \rceil$ samples. We wonder if this condition is *generically sufficient* to guarantee poisedness.

Question. *Let \mathcal{V} be a self-conjugate variety. Let $\mathcal{L}_2 \subseteq \mathbb{R}[\mathcal{V}]$ be a D -dimensional linear subspace and let $Z \subseteq \mathcal{V}$ be a generic set of samples with $|Z| \geq D/2$. Is (\mathcal{L}_2, Z) poised?*

In order to make sense of the above question, we have to be more precise about the meaning of a generic set of samples. In Section 2.2 we saw the definition of a generic sample of an irreducible variety. We have to extend this definition to multiple samples, taken possibly from a reducible variety. We now formalize the notion of genericity we use. It is slightly different from the one in Section 2.2 as it includes the complex conjugates of the samples.

Definition 4.1. Let $\mathcal{W} \subseteq \mathbb{C}^n$ be an irreducible variety and let $Z = \{z_1, \dots, z_S\} \subseteq \mathcal{W}$ be a set of samples. We say that Z satisfies a property *generically* if there is a polynomial $q \in \mathbb{C}[z_1, \dots, z_S, \overline{z_1}, \dots, \overline{z_S}]$ such that:

- $q(z_1, \dots, z_S, \overline{z_1}, \dots, \overline{z_S})$ is not identically zero for $Z \in \mathcal{W}^S$.
- the property holds whenever $q(z_1, \dots, z_S, \overline{z_1}, \dots, \overline{z_S}) \neq 0$.

Let $\mathcal{W}_1, \dots, \mathcal{W}_r$ be irreducible varieties and let $Z_1 \subseteq \mathcal{W}_1, \dots, Z_r \subseteq \mathcal{W}_r$ be sets of samples. We say that (Z_1, \dots, Z_r) satisfies a property generically if there are polynomials $q_1 \in \mathbb{C}[Z_1, \overline{Z_1}], \dots, q_r \in \mathbb{C}[Z_r, \overline{Z_r}]$ such that:

- $q_i(Z_i, \overline{Z_i})$ is not identically zero on \mathcal{W}_i , for $1 \leq i \leq r$.
- the property holds whenever $q_1(Z_1, \overline{Z_1}) \neq 0, \dots, q_r(Z_r, \overline{Z_r}) \neq 0$.

We say that Z (resp. Z_1, \dots, Z_r) is a *generic* set of samples if it satisfies generically some property of interest.

In the next section we will show that for an irreducible variety (or a conjugate pair of irreducible varieties) the answer to the question from above is positive. However, for reducible varieties, we need to make sure that we have enough samples in each irreducible component, as will be discussed in Section 4.4.

4.2. The irreducible case. Assume now that $\mathcal{V} = \mathcal{W} \cup \overline{\mathcal{W}}$, where $\mathcal{W} \subseteq \mathbb{C}^n$ is an irreducible variety. This means that either \mathcal{V} is a self-conjugate irreducible variety, or it is a conjugate pair of irreducible varieties. In the latter case, note that we can assume without loss of generality that $Z \subseteq \mathcal{W}$, by possibly exchanging some samples with their complex conjugates. We show now that if the samples Z are generic and are at least as many as the dimensionality of the problem, then the poisedness property is satisfied.

Theorem 5. *Let $\mathcal{W} \subseteq \mathbb{C}^n$ be an irreducible variety, let $\mathcal{V} = \mathcal{W} \cup \overline{\mathcal{W}}$ and let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$. Let $\mathcal{L}_2 \subseteq \mathcal{R}$ be a linear subspace and let $Z \subseteq \mathcal{W}$ be a generic set of samples. If $|Z| \geq D/2$, where D is the dimension of \mathcal{L}_2 , then (\mathcal{L}_2, Z) is poised².*

Proof. Let $v(x) \in \mathcal{R}^D$ be a basis of \mathcal{L}_2 . Let $Z_j := \{z_1, \dots, z_j\}$, let $V_j \in \mathbb{C}^{D \times j}$ be the matrix with columns $\{v(z)\}_{z \in Z_j}$ and let $\hat{V}_j := [\Re(V_j) | \Im(V_j)] \in \mathbb{R}^{D \times 2j}$. Also denote

²This theorem is an special instance of the dimensionality problem in polynomial interpolation, and more elaborate versions can be found in the literature [6].

$W_j := [\hat{V}_j | \Im(v(z_{j+1}))] \in \mathbb{R}^{D \times 2j+1}$. Because of Lemma 2, we just need to show that the matrix \hat{V}_S has rank D . To this end, we will show the following statements:

- if \hat{V}_{j-1} is full rank then W_{j-1} is generically full rank.
- if W_{j-1} is full rank then \hat{V}_j is generically full rank.

Clearly these statements imply that \hat{V}_S is full rank. Given the similarity between the two of them, we only prove the latter.

Let $j \leq D/2$ and assume that W_{j-1} is full rank. We will show that there is a polynomial $Q \in \mathbb{C}[Z_j, \overline{Z_j}]$ which is not identically zero on \mathcal{W} , and such that \hat{V}_j is full rank whenever $Q(Z_j, \overline{Z_j}) \neq 0$.

Assume that \hat{V}_j is not full rank. Then there must exist a vector $\lambda \in \mathbb{R}^{2j-1}$ such that

$$v(z_j) + v(\overline{z_j}) = 2\Re(v(z_j)) = W_{j-1}\lambda.$$

As W_{j-1} has less than D columns, there is some nonzero vector $\mu \in \mathbb{R}^D$ in its left kernel. Note that $\mu = \mu(Z_j, \overline{Z_j})$ can be parametrized as a rational function of $Z_j, \overline{Z_j}$, given that W_{j-1} is full rank. Let $q_\mu(x) := \mu^T v(x) \in \mathcal{R}$, which is nonzero due to the linear independence of $v(x)$. Observe that

$$q_\mu(z_j) + q_\mu(\overline{z_j}) = \mu^T W_{j-1} \lambda = 0.$$

As the coefficients of q_μ are rational functions on $Z_j, \overline{Z_j}$, we conclude that the samples satisfy a nonzero algebraic equation $Q \in \mathbb{C}[Z_j, \overline{Z_j}]$. \square

Remark. If the samples are real, it can be shown in a similar way that we need $|Z| \geq D$.

4.3. Verifying the number of samples. We just showed that, under genericity assumptions, the poisedness property is satisfied whenever we have as many samples as the dimension of the space. Concretely, if D is the dimension of \mathcal{L}_2 , we need to have $\lceil D/2 \rceil$ complex samples. However, as the dimension D is not known a priori, it is uncertain how many samples to take. Therefore, we need some way to estimate such dimension, and the natural quantity to consider is the empirical dimension D_e . The following corollary gives us a simple test that guarantees that $D = D_e$.

Corollary 6. *Let $\mathcal{W} \subseteq \mathbb{C}^n$ be an irreducible variety, let $\mathcal{V} = \mathcal{W} \cup \overline{\mathcal{W}}$ and let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$. Let $\mathcal{L}_2 \subseteq \mathcal{R}$ be a linear subspace and let $Z \subseteq \mathcal{W}$ be a generic set of samples. Let D be the dimension of \mathcal{L}_2 and let D_e be its empirical dimension with respect to Z . If $D_e < 2|Z|$ then (\mathcal{L}_2, Z) is poised (i.e., $D = D_e$).*

Proof. If $2|Z| < D$ it follows from the proof of Theorem 5 that $D_e = 2|Z|$. Therefore, we must have that $2|Z| \geq D$, and thus (\mathcal{L}_2, Z) is poised because of Theorem 5. \square

The above corollary suggests a simple strategy that is summarized in Algorithm 3. We form the vector $u_2(x) = \text{vec}(u(x)u(x)^T)$, whose entries span \mathcal{L}_2 . Then we build the matrix of evaluations \hat{U}_2 with columns $u_2(z)$ for $z \in Z \cup \overline{Z}$. The rank of this matrix is the empirical dimension D_e . Thus, if \hat{U}_2 does not have full column rank the above corollary holds.

Algorithm 3 Test samples**Input:** Polynomial vector $u(x)$, samples Z of a variety \mathcal{V} **Output:** “True”, if generically we must have that (\mathcal{L}_2, Z) is poised, where $\mathcal{L}_2 \subseteq \mathbb{R}[\mathcal{V}]$ is spanned by $u(x)u(x)^T$. “False”, if we cannot guarantee it.

```

1: procedure GOODSAMPLES( $u(x), Z$ )
2:    $\hat{U}_2 :=$  matrix with columns  $\text{vec}(u(z)u(z)^T)$ , for  $z \in Z \cup \bar{Z}$ 
3:   if  $\hat{U}_2$  has full column rank then
4:     return False
5:   return True

```

Remark. Consider the Hermitian matrix $\hat{U}_2^* \hat{U}_2$, where $*$ denotes the conjugate transpose. This matrix is often much smaller than \hat{U}_2 , and it can be constructed efficiently as

$$\hat{U}_2^* \hat{U}_2 = [\langle u(z_i), u(z_j) \rangle^2]_{z_i, z_j \in Z \cup \bar{Z}} = (\hat{U}^* \hat{U}) \circ (\hat{U}^* \hat{U})$$

where \circ denotes the Hadamard product. Therefore, it is practical to use matrix $\hat{U}_2^* \hat{U}_2$ instead of \hat{U}_2 , given that they have the same rank.

Example 4.1. Consider the case of Example 3.2. We used $S = 3$ samples to compute the S-SOS pre-certificate. To verify that the number of samples was sufficient, we construct the matrix

$$\hat{U}_2^* \hat{U}_2 = \begin{bmatrix} 1.5581 & -0.2937+0.2562i & 0.1730-0.1158i & 0.0902+0.1118i & 0.0981 & -0.0676-0.0720i \\ -0.2937-0.2562i & 1.5581 & 0.1730+0.1158i & 0.0981 & 0.0902-0.1118i & -0.0676+0.0720i \\ 0.1730+0.1158i & 0.1730-0.1158i & 0.2535 & -0.0676-0.0720i & -0.0676+0.0720i & 0.1396 \\ 0.0902-0.1118i & 0.0981 & -0.0676+0.0720i & 1.5581 & -0.2937-0.2562i & 0.1730+0.1158i \\ 0.0981 & 0.0902+0.1118i & -0.0676-0.0720i & -0.2937+0.2562i & 1.5581 & 0.1730-0.1158i \\ -0.0676+0.0720i & -0.0676-0.0720i & 0.1396 & 0.1730-0.1158i & 0.1730+0.1158i & 0.2535 \end{bmatrix}$$

The rank of this matrix is 5, and thus the condition from Corollary 6 is satisfied. Therefore, the number of samples is sufficient.

4.4. Reducible varieties. The analysis made so far makes an irreducibility assumption on the variety \mathcal{V} . This assumption is satisfied for many varieties, in particular for any variety parametrized by \mathbb{C}^n . Even if \mathcal{V} is not irreducible, we can always work with each of its irreducible components independently. Indeed, note that $p \geq 0$ on some variety if and only if $p \geq 0$ on each irreducible component.

Nonetheless, there are circumstances in which we may not want to impose an irreducibility assumption. For example, if the variety has bad numerical properties and thus its irreducible components cannot be accurately estimated. In such situations, we can repeat the same analysis from before if we have some method that samples points from each irreducible component. For instance, if we intersect the variety \mathcal{V} with a generic hyperplane of complementary dimension, the intersection is a finite set that contains points in each irreducible component. Note that we do not know which component do the samples belong to, but we are certain that there is at least one sample in each component.

The following corollary shows that if we have a sample set with enough points on each irreducible component, then (\mathcal{L}_2, Z) is poised.

Corollary 7. *Let $\mathcal{W} \subseteq \mathbb{C}^n$ be a variety, let $\mathcal{V} = \mathcal{W} \cup \bar{\mathcal{W}}$ and let $\mathcal{R} = \mathbb{R}[\mathcal{V}]$. Let $\mathcal{L}_2 \subseteq \mathcal{R}$ be a linear subspace. Let $\mathcal{W} = \mathcal{W}_1 \cup \dots \cup \mathcal{W}_r$ be the irreducible decomposition, and let*

$Z_1 \subseteq \mathcal{W}_1, \dots, Z_r \subseteq \mathcal{W}_r$ be generic sets of samples. If $|Z_i| \geq D/2$ for all i , where D is the dimension of \mathcal{L}_2 , then (\mathcal{L}_2, Z) is poised.

Proof. Let $f \in \mathcal{L}_2$ be such that $f(z) = 0$ for all $z \in Z$. We want to show that f is the zero polynomial in $\mathbb{C}[\mathcal{V}]$. Let $\mathcal{V}_i := \mathcal{W}_i \cup \overline{\mathcal{W}_i}$ and let $\psi_i : \mathbb{C}[\mathcal{V}] \rightarrow \mathbb{C}[\mathcal{V}_i]$ be the restriction operator. It is clear that the dimension of $\psi_i(\mathcal{L}_2)$ is at most D . Thus, Theorem 5 says that $(\psi_i(\mathcal{L}_2), Z_i)$ is poised whenever $q_i(Z_i, \overline{Z_i}) \neq 0$, for some polynomial q_i which is nonzero on \mathcal{W}_i . Note that $\psi_i(f)$ evaluates to zero on Z_i , and thus $\psi_i(f)$ must be the zero element in $\mathbb{C}[\mathcal{V}_i]$ whenever $q_i(Z_i, \overline{Z_i}) \neq 0$. Finally, observe that $(\psi_1 \times \dots \times \psi_k) : \mathbb{C}[\mathcal{V}] \rightarrow \mathbb{C}[\mathcal{V}_1] \times \dots \times \mathbb{C}[\mathcal{V}_k]$ is injective. We conclude that whenever $q_i(Z_i, \overline{Z_i}) \neq 0$ then $(\psi_1 \times \dots \times \psi_k)(f)$ is zero and thus f must be zero. \square

4.5. Computing S-SOS certificates. We already developed all the tools needed to compute an S-SOS certificate, and we now put them together. Given a polynomial p and an irreducible variety \mathcal{V} the sampling SOS methodology is as follows.

- (I) Let d be a degree bound and S be a guess on the number of samples. Note that $\binom{n+2d}{2d}$ samples are always sufficient, but possibly a much smaller S will work.
- (II) Let $u(x)$ be the vector with all monomials up to degree d , and let $Z \subseteq \mathcal{V}$ be a generic set of S samples.
- (III) Orthogonalize $u(x)$ using Algorithm 2.
- (IV) Check if the number of samples is sufficient with Algorithm 3. If insufficient, increase S (e.g., multiplicative update).
- (V) Compute an S-SOS pre-certificate with the SDP in (7). If the SDP is infeasible, the degree bound d must be increased.
- (VI) Verify that the pre-certificate is correct using Algorithm 1. If incorrect, then must likely there was a numerical problem.

Naturally, the most computationally expensive part is solving the SDP. Note that for reducible varieties we have to take into account the considerations from Section 4.4.

5. EXAMPLES

We now show several examples and numerical evaluations to illustrate our methodology. We implemented our methods in Matlab, using SDPT3 [25] to solve the semidefinite programs. We also use Bertini [1] to compute numerical irreducible decompositions and Macaulay2 [9] for Gröbner bases. The experiments are performed on an i7 PC with 8 cores of 3.40 GHz, 15.6 GB RAM, running Ubuntu 14.04.

We will compare our techniques with the following traditional SOS methodologies.

Standard SDP: The simplest way to certify nonnegativity is by finding equations

F, g_1, \dots, g_m satisfying (2). This reduces to an SDP, as seen in Appendix A.1.

Gröbner bases: There is a natural quotient ring formulation based on a Gröbner basis of h . We explain this approach in Appendix A.2.

We recall that SOS programs are often used for polynomial optimization problems. Indeed, minimizing a polynomial $p \in \mathbb{R}[x]$ is equivalent to finding the largest γ such that $p(x) - \gamma$ is nonnegative, which is relaxed to be an SOS polynomial. Some of the examples below are SOS relaxations of an optimization problem.

The solution γ^* obtained with the SDP gives a valid lower bound on the minimum of $p(x)$, and it is often the case that it is the true minimum. Moreover, if the minimizer is unique and the dual matrix is rank one, then the minimizer x^* might be recovered from the dual variables. In particular, $x^* = \Re(\sum_s y_s^* z_s)$ for the sampling SDP in (7), where y_s^* are the (complex) dual variables of the equality constraints. For a review of SOS methods for optimization we refer to [4, 13].

5.1. Nilpotent matrices. Let $\mathcal{V} := \{X \in \mathbb{C}^{n \times n} : X^n = 0\}$ be the variety of nilpotent matrices. Let $p(X) := \det(X + \text{id}_n)$, which is nonnegative on \mathcal{V} (moreover, it is identically one). We compare different SOS methodologies to certify this.

First, consider the sampling approach. It turns out that any degree bound d will work, so assume that $d = 1$. Let's take $S = \binom{n+d}{d} = n+1$ samples, which are always sufficient. Note that it is very easy to sample nilpotent matrices. For instance, we can generate a random triangular matrix with zero diagonal, and then apply a similarity transformation. For each sample $X_s \in \mathcal{V}$, we can efficiently evaluate $p(X_s)$ by using Gaussian elimination. As $p(X_s) = 1$ for all samples X_s , we will obtain the trivial SOS decomposition $p(X) = 1^2$.

Consider now the Gröbner bases approach. Let $h \subseteq \mathbb{R}[X]$ be the n^2 equations given by $X^n = 0$. We want to compute a Gröbner basis of h . Note, however, that the total number of terms in h is on the order of n^{n+1} , and the polynomials are all of degree n . Therefore, this Gröbner basis computation is extremely complicated.

If we are smarter, we can take a different set of defining equations of \mathcal{V} . Consider the polynomial $Q_X(t) := \det(t \text{id}_n - X) - t^n$, and let $h' \subseteq \mathbb{R}[X]$ be the equations given by the coefficients of $Q_X(t)$. It turns out that h' also defines the variety \mathcal{V} and furthermore, $\langle h' \rangle$ is the radical ideal of $\langle h \rangle$ [10, §7]. In addition, it can be seen that h' is already a Gröbner basis. Note, however, that the number of terms of $Q_X(t)$, and thus of h' , is on the order of $n!$.

Once we have the Gröbner basis h' , we need to compute the normal form of p . To obtain this normal form we need to consider p as a dense polynomial. As both p and h' have on the order of $n!$ terms, performing this reduction is computationally intractable. If we are able to reduce it, we will conclude that $p(X) = 1$, as before.

Finally, note that the standard SDP methodology suffers from the same problems of the Gröbner bases approach. For this method there is an additional problem, which is that the monomial basis $u(X)$ will be very large in order to account for all the monomials in $p(X)$ and $h(X)$. This problem was avoided in the previous methods because of the quotient ring reductions.

This example illustrates two of the advantages of the sampling SOS formulation: it avoids the algebraic problem of deciding which equations to use (e.g., h vs. h'), and it allows the use of straight-line programs (e.g., Gaussian elimination) for more efficient evaluations. Note that the variety equations can also be given as straight-line programs by taking advantage of numerical algebraic geometry.

5.2. Weighted orthogonal Procrustes. We consider a family of optimization problems over varieties of orthogonal matrices. The Stiefel manifold $St(k, \mathbb{R}^n)$ is the set of orthonormal k -frames in \mathbb{R}^n . We identify it with the set of matrices $X \in \mathbb{R}^{n \times k}$ such

that $X^T X = \text{id}_k$. Note that we can easily sample points from this variety, for instance, by using the Cayley parametrization. Alternatively, we can orthogonalize a random real matrix.

The weighted orthogonal Procrustes problem, also known as Penrose regression problem, asks for a matrix $X \in St(k, \mathbb{R}^n)$ that minimizes $\|AXC - B\|$, for some matrices $A \in \mathbb{R}^{m_1 \times n}$, $B \in \mathbb{R}^{m_1 \times m_2}$, $C \in \mathbb{R}^{k \times m_2}$. There is no closed form solution for this problem, and several local optima may exist [5, 27].

Given a bound d , we consider the vector $u(x)$ with all monomials of degree at most d . The sampling SDP formulation of the degree d SOS relaxation is:

$$\begin{aligned} & \max_{\gamma \in \mathbb{R}, Q \succeq 0} \quad \gamma \\ & \text{subject to} \quad \|AX_s C - B\|^2 - \gamma = Q \bullet u(X_s)u(X_s)^T, \quad \text{for } s = 1, \dots, S \\ & \quad \quad \quad X_s \in St(k, \mathbb{R}^n) \end{aligned}$$

Example 5.1 ([5], Ex 2). Assume that C is the 3×3 identity matrix and

$$A^T = \begin{bmatrix} 0.2190 & 0.0470 & 0.6789 & 0.6793 & 0.9347 \\ 0.3835 & 0.5194 & 0.8310 & 0.0346 & 0.0535 \\ 0.5297 & 0.6711 & 0.0077 & 0.3834 & 0.0668 \\ 0.4175 & 0.6868 & 0.5890 & 0.9304 & 0.8462 \end{bmatrix}, \quad B^T = \begin{bmatrix} 0.6526 & 0.2110 & 0.2229 & -0.4104 & -0.9381 \\ 0.6942 & 0.2204 & 0.2015 & 0.2994 & 1.0943 \\ 0.8299 & 1.1734 & -0.1727 & 0.0474 & -0.2351 \end{bmatrix}.$$

We consider the degree 1 SOS relaxation. Solving the above SDP we obtain a lower bound of 1.118147 on the minimum norm $\|AXC - B\|$. Furthermore, the dual SDP matrix has rank one, and thus we can obtain a solution achieving such lower bound:

$$(X^*)^T = \begin{bmatrix} -0.0895 & 0.7472 & 0.2732 & -0.5992 \\ 0.7726 & -0.1843 & 0.6035 & -0.0702 \\ -0.5277 & 0.0163 & 0.7309 & 0.4324 \end{bmatrix}.$$

Table 1 compares different SDP formulations of the degree 1 SOS relaxation of the weighted Procrustes problem. We consider the case where $m_1 = n$ and $m_2 = r$. The table shows the number of variables/constraints and the computation time for the standard SDP and the sampling SDP. The computation is performed on random instances, in which matrices A , B , C are generated from the standard normal distribution. For the sampling SDP we use the image form of the SDP (see Section 3.1), given that it has low codimension. We remark that for the sampling SDP we include the preprocessing time, i.e., Algorithms 2 and 3.

Table 1 also shows the computation time of a Gröbner basis. Note that Macaulay2 ran out of memory starting on $n = 7$, $k = 5$. As expected, computing Gröbner bases is prohibitive for relatively small values of n, k .

5.3. Trace ratio problem. We consider now a problem over the Grassmannian manifold $Gr(k, \mathbb{R}^n)$, which is the set of all k -dimensional subspaces of \mathbb{R}^n . By identifying a subspace with the orthogonal projection matrix onto it, we can view $Gr(k, \mathbb{R}^n)$ as the set of symmetric matrices $X \in \mathbb{R}^{n \times n}$ satisfying $X^2 = X$ and $\text{tr}(X) = k$. The trace ratio problem looks for the maximizer of $\frac{\text{tr}(AX)}{\text{tr}(BX)}$ on $Gr(k, \mathbb{R}^n)$, for some given matrices $A, B \in \mathbb{R}^{n \times n}$, $B \succ 0$. This problem arises in machine learning, and it can be efficiently solved by iterative methods, given that it has a unique local maximum [28]. We consider the following variation:

$$\max_{X \in Gr(k, \mathbb{R}^n)} \frac{\text{tr}(AX)}{\text{tr}(BX)} + \text{tr}(CX)$$

TABLE 1. Degree 1 SOS relaxations for the weighted Procrustes problem

n	r	Standard SDP			Sampling SDP			Gröbner bases time(s)
		variables	constraints	time(s)	variables	constraints	time(s)	
4	2	178	73	0.52	46	42	0.10	0.00
5	3	682	233	0.65	137	130	0.11	0.03
6	4	1970	576	1.18	326	315	0.15	9.94
7	5	4727	1207	3.56	667	651	0.31	out of mem.
8	6	9954	2255	13.88	1226	1204	0.70	out of mem.
9	7	19028	3873	42.14	2081	2052	2.11	out of mem.
10	8	33762	6238	124.43	3322	3285	5.07	out of mem.

for some $A, B, C \in \mathbb{R}^{n \times n}$, $B \succ 0$. This problem may have several local maxima and thus local methods may not converge to the global optimum [29, 30].

To obtain an SOS relaxation, note that the problem is equivalent to minimizing γ such that $\text{tr}(BX)(\gamma - \text{tr}(CX)) - \text{tr}(AX)$ is nonnegative on $Gr(k, \mathbb{R}^n)$. Thus, the SDP to consider is:

$$\begin{aligned} \min_{\gamma \in \mathbb{R}, Q \succeq 0} \quad & \gamma \\ \text{subject to} \quad & \text{tr}(BX_s)(\gamma - \text{tr}(CX_s)) - \text{tr}(AX_s) = Q \bullet u(X_s)u(X_s)^T, \quad \text{for } s = 1, \dots, S \\ & X_s \in Gr(k, \mathbb{R}^n) \end{aligned}$$

Example 5.2 ([30], Ex 3.1). Consider the problem for the matrices A, B, C from below. Using the degree 1 SOS relaxation we obtain an upper bound of 28.692472. As the dual matrix has rank one, we can recover the optimal solution X^* .

$$A = \begin{bmatrix} 11 & 5 & 8 \\ 5 & 10 & 9 \\ 8 & 9 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & 7 & 7 \\ 7 & 10 & 8 \\ 7 & 8 & 8 \end{bmatrix}, \quad C = \begin{bmatrix} 15 & 10 & 9 \\ 10 & 7 & 6 \\ 9 & 6 & 6 \end{bmatrix}, \quad X^* = \begin{bmatrix} 0.61574 & 0.15424 & 0.46132 \\ 0.15424 & 0.93809 & -0.18517 \\ 0.46132 & -0.18517 & 0.44617 \end{bmatrix}$$

As before, we compare the standard SDP and the sampling SDP of the degree 1 SOS relaxation. Table 2 shows the number of variables/constraints and the computation time on random instances for both methods. It also shows the computation time of a Gröbner basis.

TABLE 2. Degree 1 SOS relaxations for the trace ratio problem

n	r	Standard SDP			Sampling SDP			Gröbner bases time(s)
		variables	constraints	time(s)	variables	constraints	time(s)	
4	2	342	188	0.47	56	45	0.10	0.00
5	3	897	393	0.71	121	105	0.11	0.02
6	4	2062	738	1.34	232	210	0.15	0.20
7	5	4265	1277	3.62	407	378	0.19	6.04
8	6	8106	2073	9.06	667	630	0.34	488.17
9	7	14387	3198	23.83	1036	990	0.61	out of mem.
10	8	24142	4733	58.17	1541	1485	1.18	out of mem.

5.4. Low rank approximation. Consider the problem of finding the nearest rank k tensor. Let $\mathbb{C}^{n_1 \times \dots \times n_\ell}$ denote the set of tensors of order ℓ and dimensions (n_1, \dots, n_ℓ) and let $\mathbb{C}_{\leq k}^{n_1 \times \dots \times n_\ell}$ be the closure of the space of tensors of rank at most k . Note that we can easily generate generic samples of rank k tensors. Given a real tensor $T \in \mathbb{R}^{n_1 \times \dots \times n_\ell}$, the rank k approximation problem asks for the nearest point $X \in \mathbb{R}_{\leq k}^{n_1 \times \dots \times n_\ell}$, i.e., the minimizer of the squared norm $\|T - X\|^2 := \sum_{i_1, \dots, i_\ell} (T_{i_1 \dots i_\ell} - X_{i_1 \dots i_\ell})^2$.

For a tensor X consider the SOS polynomial $\varsigma(X) := \sum_{i_1, \dots, i_\ell} X_{i_1 \dots i_\ell}^2$. Let $d := \lfloor k/2 \rfloor + 1$ and let $u(X)$ be the vector with all monomials of degree at most d . We consider the following SDP relaxation:

$$\begin{aligned} & \max_{\gamma \in \mathbb{R}, Q \succeq 0} \quad \gamma \\ & \text{subject to} \quad (\|T - X_s\|^2 - \gamma) \varsigma(X_s)^{d-1} = Q \bullet u(X_s)u(X_s)^T, \quad \text{for } s = 1, \dots, S \\ & \quad \quad \quad X_s \in \mathbb{C}_{\leq k}^{n_1 \times \dots \times n_\ell} \end{aligned}$$

We remark that computing the defining equations of the variety $\mathbb{C}_{\leq k}^{n_1 \times \dots \times n_\ell}$ is very complicated [12]. This means that using the traditional SOS methodologies is usually not possible.

Example 5.3 ([7], Ex 3). Let $T \in \mathbb{R}^{2 \times 2 \times 2 \times 2}$ be the tensor whose nonzero entries are

$$T_{1111} = 25.1, \quad T_{1121} = 0.3, \quad T_{1212} = 25.6, \quad T_{2111} = 0.3, \quad T_{2121} = 24.8, \quad T_{2222} = 23.$$

Consider the rank one approximation problem, using the SDP relaxation above. Based on Corollary 6, we require 49 generic samples on $\mathbb{C}_{\leq 1}^{2 \times 2 \times 2 \times 2}$. Solving the SDP we obtain the lower bound 42.1216 on the minimum distance $\|T - X\|$. From the dual solution we recover the minimizer X^* , whose only nonzero entry is $X_{1212}^* = 25.6$.

Consider now the rank three approximation problem. We require 2422 generic samples on $\mathbb{C}_{\leq 3}^{2 \times 2 \times 2 \times 2}$. The SDP gets a lower bound of 23.0000 on the minimum distance. Again, we can recover the minimizer X^* from the dual:

$$X_{1111}^* = 25.1, \quad X_{1121}^* = 0.3, \quad X_{1212}^* = 25.6, \quad X_{2111}^* = 0.3, \quad X_{2121}^* = 24.8$$

where the remaining entries are zero. To see that X^* is indeed rank three, note that after removing the entry 25.6 we are left with a 2×2 matrix.

5.5. Certifying infeasibility. Given a complex variety $\mathcal{V} \subseteq \mathbb{C}^n$ consider the problem of certifying that $\mathcal{V} \cap \mathbb{R}^n$ is empty. A *Positivstellensatz* infeasibility certificate consists in showing that the constant polynomial -1 is SOS on the variety \mathcal{V} [16]. For instance, if $\mathcal{V} = \{i, -i\} \subseteq \mathbb{C}$, a Positivstellensatz certificate is that $-1 = x^2$ on the variety \mathcal{V} . We take an approach of numerical algebraic geometry, where we first compute a witness set for each irreducible component of \mathcal{V} , and then use the sampling SOS formulation to obtain the infeasibility certificate. For a given vector $u(x)$ the SDP problem to solve is:

$$\begin{aligned} & \text{find} \quad Q \succeq 0 \\ & \text{subject to} \quad -1 = Q \bullet u(z_s)u(z_s)^T, \quad \text{for } s = 1, \dots, S \\ & \quad \quad \quad z_s \in \mathcal{V} \end{aligned}$$

Example 5.4. Let $\mathcal{V} \subseteq \mathbb{C}^9$ be the positive dimensional part of the cyclic 9-roots problem. The cyclic 9-roots equations are:

$$\begin{aligned}
& x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 \\
& x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 + x_5x_6 + x_6x_7 + x_7x_8 + x_8x_9 + x_9x_1 \\
& x_1x_2x_3 + x_2x_3x_4 + x_3x_4x_5 + x_4x_5x_6 + x_5x_6x_7 + x_6x_7x_8 + x_7x_8x_9 + x_8x_9x_1 + x_9x_1x_2 \\
& \vdots \\
& x_1x_2x_3x_4x_5x_6x_7x_8 + x_2x_3x_4x_5x_6x_7x_8x_9 + \cdots + x_9x_1x_2x_3x_4x_5x_6x_7 \\
& x_1x_2x_3x_4x_5x_6x_7x_8x_9 - 1
\end{aligned}$$

The zero set of these equations consists of a two-dimensional variety \mathcal{V} of degree 18, and 6156 isolated solutions [8]. We remark that computing a Gröbner basis of these equations is very complicated unless its special structure is exploited. In particular, the command `gb` of Macaulay2 ran out of memory after 5 hours of computation.

We found a numerical irreducible decomposition of \mathcal{V} using Bertini. It took 2h 45m to obtain such decomposition. The variety \mathcal{V} decomposes into six irreducible varieties of degree 3. For each of the six components we consider the sampling formulation of the degree 2 SOS relaxation. Based on Corollary 6, we require 31 complex samples on each component. For each $j = 0, \dots, 5$ we solved the respective sampling SDP, obtaining an infeasibility certificate of the form

$$-1 = (R_j u(x))^T (R_j u(x)), \quad \text{for } x \in \mathcal{V}_j.$$

This allows us to conclude that each irreducible component of \mathcal{V} is purely complex. For instance, for the first irreducible component \mathcal{V}_0 it takes only 0.74s to obtain the certificate shown in Figure 1.

5.6. Amoeba membership. The (unlog) amoeba $\mathcal{A}_{\mathcal{V}} \subseteq \mathbb{R}_+^n$ of a variety $\mathcal{V} \subseteq \mathbb{C}^n$ is the image of \mathcal{V} under the absolute value function, i.e., $\mathcal{A}_{\mathcal{V}} = \{|z| : z \in \mathcal{V}\}$. The amoeba membership problem is to determine whether some point $\lambda \in \mathbb{R}_+^n$ belongs to $\mathcal{A}_{\mathcal{V}}$. Theobald and De Wolff recently proposed the use of Positivstellensatz certificates to prove that $\lambda \notin \mathcal{A}_{\mathcal{V}}$ [24]. We now briefly describe this approach.

For some $f \in \mathbb{C}[z]$, let $\Re(f), \Im(f) \in \mathbb{R}[x, y]$ be such that

$$f(x + iy) = \Re(f)(x, y) + i \Im(f)(x, y).$$

Consider the following sets of equations in $\mathbb{R}[x, y]$:

$$J_{\mathcal{V}} := \{\Re(f_j), \Im(f_j)\}_{j=1}^m, \quad h_{\lambda} := \{x_i^2 + y_i^2 - \lambda_i^2\}_{i=1}^n$$

where f_j are the defining equations of \mathcal{V} . Theobald and De Wolff suggest computing a Gröbner basis of $J_{\mathcal{V}} \cup h_{\lambda}$ and then search for a Positivstellensatz infeasibility certificate.

Consider the following approach based on a set of samples $Z \subseteq \mathcal{V}$. Let $\hat{\mathcal{V}} \in \mathbb{C}^{2n}$ be the zero set of $J_{\mathcal{V}} \subseteq \mathbb{R}[x, y]$. Note that if $z \in \mathcal{V}$ then $(\Re(z), \Im(z)) \in \hat{\mathcal{V}}$. Thus, given some monomial vectors $u(x, y)$ and $v(x, y)$, we can formulate the following SDP:

$$\begin{aligned}
& \text{find} \quad Q \succeq 0, \ C \\
& \text{subject to} \quad -1 = Q \bullet u(x_s, y_s)u(x_s, y_s)^T + h_{\lambda}(x_s, y_s)^T C v(x_s, y_s), \quad \text{for } s = 1, \dots, S \\
& \quad z_s = x_s + iy_s \in \mathcal{V}
\end{aligned}$$

$$u(x) = (x_8^2, x_7x_9, x_6^2, x_5x_9, x_5x_7, x_4^2, x_3x_6, x_2x_7, x_2x_6, x_2^2, x_1x_3, x_1^2, x_8, x_7, x_6, x_5, x_3, x_1, 1)$$

$$R_0 = \begin{bmatrix} -0.9638686 & -0.3445318 & 0.8395791 & -1.9531033 & 0.6329543 & -0.0152284 & 0.0238164 & 0.4701138 & -1.9766327 & -0.8363703 \\ 0.3474835 & -0.3993919 & 0.5501348 & -1.2198730 & -0.2314149 & 0.0354563 & 1.0086575 & 0.4018444 & 1.0316339 & -0.6193326 \\ 0.0117704 & -0.5278490 & 0.6157589 & -0.3131173 & 0.2207819 & -0.0080541 & 0.4038186 & 0.1500184 & -0.2618475 & 0.3089739 \\ -0.0131866 & 0.1597228 & 0.1191077 & -0.1088218 & 0.0697348 & -0.1149430 & -0.5067092 & -0.1883695 & -0.5993569 & 0.0244521 \\ -0.4504113 & -0.0761266 & 0.0056933 & 0.1535964 & -0.0860039 & 0.0007534 & 0.1264270 & 0.0880389 & -0.0927822 & -0.1429983 \\ -0.0804265 & 0.1450405 & -0.0077285 & -0.1657304 & -0.3240087 & 0.0014097 & 0.0631496 & -0.4083965 & 0.0191162 & 0.0854950 \\ 0.0192110 & 0.1019831 & -0.1208989 & -0.1821975 & -0.1203214 & 0.0405222 & 0.0595267 & -0.1921851 & -0.0669972 & -0.1710978 \\ -0.1242984 & 0.1450764 & -0.2725352 & -0.1145423 & 0.0498037 & 0.0036466 & -0.0705293 & 0.2012444 & 0.0873671 & -0.2016367 \\ -0.0724047 & -0.0072012 & -0.0659910 & 0.1174698 & -0.0830511 & 0.0339559 & 0.1872153 & -0.0010648 & -0.1488432 & 0.1014557 \\ -0.1440253 & 0.0026597 & -0.1198142 & 0.0147434 & 0.1104305 & 0.0249783 & 0.0246079 & -0.0286915 & -0.0633959 & 0.0786306 \\ 0.0872131 & -0.0503333 & -0.0426310 & -0.0108485 & -0.1538320 & 0.0351373 & -0.0895051 & 0.0994606 & -0.0858176 & 0.0033518 \\ 0.0508126 & 0.0850471 & -0.0581353 & -0.0654513 & 0.0413446 & -0.0119532 & 0.0757765 & 0.0459333 & -0.0169990 & 0.1009677 \\ 0.0112157 & 0.0251923 & -0.0096306 & -0.0680984 & -0.0005761 & 0.0111481 & -0.0373533 & 0.0293210 & 0.0134771 & 0.1119457 \\ 0.0240035 & -0.1089833 & -0.0750114 & -0.0316807 & 0.0593823 & -0.0045386 & -0.0385441 & -0.0541272 & 0.0162211 & 0.0093965 \\ 0.0727416 & 0.0067146 & -0.0258618 & 0.0206007 & 0.0284529 & -0.0125337 & 0.0450957 & -0.0142651 & -0.0460162 & -0.0377493 \\ 0.0018262 & 0.0096039 & 0.0092749 & -0.0098153 & 0.0116513 & 0.0124708 & -0.0166840 & -0.0307406 & 0.0039079 & -0.0005090 \\ 0.0167276 & 0.0418916 & 0.0339853 & 0.0127189 & 0.0353915 & 0.0352643 & -0.0230590 & 0.0037635 & -0.0020096 & -0.0118329 \\ -0.0000118 & 0.0028062 & 0.0010184 & -0.0000054 & 0.0000008 & -0.0076840 & -0.0004971 & 0.0000175 & 0.0000017 & 0.0000097 \\ 0.0000403 & 0.0016845 & -0.0003547 & -0.0000443 & -0.0000077 & 0.0102089 & 0.0023837 & -0.0001760 & -0.0000343 & -0.0000479 \\ 0.1130541 & 0.0243746 & -1.0601901 & -0.5184653 & 0.5389394 & -0.5290480 & 1.4666654 & -0.3021666 & -0.0722647 & \\ -0.4838530 & 0.0446110 & 0.0443714 & 0.0400056 & -1.4678116 & -0.8155807 & -0.3859305 & 0.4715178 & 0.0326426 & \\ -0.1549109 & 0.0903295 & 0.6966522 & 0.1005015 & 0.1763720 & 1.0574739 & -0.3501351 & -0.0462028 & 0.1140547 & \\ 0.5739993 & -0.0684158 & 0.3571626 & 0.0861604 & -0.6655387 & -0.1886137 & -0.4910517 & 0.1000489 & 0.1425230 & \\ -0.0397509 & 0.0175350 & -0.4837186 & 0.1313369 & 0.0071916 & 0.1063667 & -0.5799628 & 0.0186656 & -0.1366172 & \\ 0.0288713 & -0.0270858 & -0.1711768 & 0.0516328 & -0.2256508 & 0.3277098 & 0.1917983 & -0.0453268 & -0.0388412 & \\ -0.1128892 & 0.0088688 & 0.1873327 & 0.1850600 & 0.2445545 & -0.1280363 & -0.1585725 & -0.1034274 & 0.0225831 & \\ 0.0175299 & -0.0263491 & 0.0802817 & -0.1067258 & -0.0673210 & 0.2239093 & 0.0003892 & -0.0262654 & 0.1438365 & \\ -0.1171573 & 0.0113433 & 0.0483356 & -0.1727556 & -0.1006414 & -0.0966047 & -0.0033401 & -0.1640771 & 0.2015432 & \\ -0.1481463 & 0.0152680 & 0.0863339 & 0.1057026 & -0.0780024 & -0.0238758 & 0.0753908 & 0.1662874 & -0.0984320 & \\ -0.0868955 & 0.0296646 & -0.0024721 & -0.0139792 & -0.0210549 & 0.0241345 & -0.0051360 & 0.0220213 & -0.0734852 & \\ 0.0428166 & -0.0263798 & -0.0151096 & -0.0341225 & 0.0015975 & -0.0045648 & -0.0459362 & -0.0584079 & -0.1230225 & \\ -0.0239812 & -0.0019790 & 0.0630874 & 0.0477949 & 0.0312372 & -0.0118692 & -0.0260705 & 0.0443067 & 0.1156214 & \\ -0.0084732 & 0.0434794 & -0.0322293 & -0.0031909 & -0.0252367 & 0.0031306 & -0.0173622 & -0.0663103 & -0.0011153 & \\ -0.0013241 & -0.0085007 & 0.0350008 & 0.0115394 & -0.0048681 & 0.0243468 & -0.0006443 & 0.0436652 & 0.0255582 & \\ -0.0261973 & -0.0090256 & -0.0019749 & -0.0604685 & 0.0112659 & 0.0022552 & -0.0232959 & 0.0276527 & -0.0063107 & \\ -0.0469611 & 0.0047957 & -0.0132026 & 0.0243877 & -0.0253729 & 0.0101839 & -0.0006742 & -0.0451327 & -0.0022650 & \\ -0.0020972 & 0.0071188 & -0.0000255 & -0.0008521 & 0.0000673 & 0.0000274 & -0.0002656 & 0.0000749 & -0.0000434 & \\ 0.0054816 & 0.0120243 & 0.0000808 & -0.0010943 & 0.0006066 & 0.0000603 & 0.0000083 & 0.0026580 & 0.0001099 & \end{bmatrix}$$

FIGURE 1. Positivstellensatz infeasibility certificate for the cyclic 9-roots problem.

Example 5.5. Let $\mathcal{V} \subseteq \mathbb{C}^{nk}$ be the complex variety associated to the Stiefel manifold $St(k, \mathbb{R}^n)$. Let $\lambda = (1/n, 1/n, \dots, 1/n)$, and let's show that $\lambda \notin \mathcal{A}_{\mathcal{V}}$ using the SDP from above. We consider the degree 1 SOS relaxation for the case $n = 6, k = 4$. We require 1205 complex samples on \mathcal{V} , which we obtain using the Cayley parametrization. It takes only 0.79s to compute the Positivstellensatz certificate from below. On the other hand, Macaulay2 ran out of memory while computing a Gröbner basis of $J_{\mathcal{V}}$.

$$-1 = (Ru(x, y))^T (Ru(x, y)) - 1.2 \sum_{i=1}^6 h_i(x, y), \quad \text{for } (x, y) \in \hat{\mathcal{V}}$$

$$u(x, y) = (y_6, y_5, y_4, y_3, y_2, y_1) \quad h_i(x, y) = x_i^2 + y_i^2 - 1/n^2$$

$$R = \begin{bmatrix} 0.1765714 & 0.8458754 & -0.3371163 & -1.0598462 & 0.0269367 & 0.6447252 \\ 0.2893688 & 0.1328983 & -1.4142041 & 0.4346374 & 0.1677938 & -0.2855976 \\ -0.4505154 & -0.6521358 & -0.3240160 & 0.2748310 & -0.0022626 & 1.2614402 \\ 1.0819066 & 0.4199281 & 0.3317461 & 0.7231132 & -0.3725210 & 0.5304889 \\ 0.8377745 & -1.0150421 & -0.1600336 & -0.6991182 & -0.3744590 & -0.1150085 \\ 0.4579696 & -0.1868200 & 0.2138378 & -0.0250102 & 1.4464173 & 0.1299494 \end{bmatrix}$$

APPENDIX A. TRADITIONAL SOS CERTIFICATES

This section reviews two common SOS methodologies to certify nonnegativity on a variety, comparing them with our sampling formulation.

A.1. Standard SDP formulation. Let \mathcal{V} be a variety with defining equations $h = (h_1, \dots, h_m)$, and let $p \in \mathbb{R}[x]$ be nonnegative on $\mathcal{V} \cap \mathbb{R}^n$. The standard approach to certify this nonnegativity is to compute an SOS polynomial F and a tuple of polynomials $g = (g_1, \dots, g_m)$ that satisfy equation (2). Let $u(x) \in \mathbb{R}[x]^N$ be a given monomial vector,

typically consisting of all monomials up to certain degree bound. Then computing an SOS certificate (F, g) reduces to the following problem:

$$(9) \quad \begin{aligned} & \text{find} \quad Q \in \mathbb{R}^{N \times N}, \quad C \in \mathbb{R}^{m \times N}, \quad Q \succeq 0 \\ & \text{subject to} \quad p(x) = Q \bullet u(x)u(x)^T + h(x)^T C u(x) \end{aligned}$$

where $F(x) = Q \bullet u(x)u(x)^T$ and $g(x) = C u(x)$. Note that the polynomial equality above is an affine relation in the entries of Q and C , and thus it is indeed an SDP.

Example A.1. Let's retake the case from Example 3.2, i.e., we want to certify that $p(X) = 4X_{21} - 2X_{11}X_{22} - 2X_{12}X_{21} + 3$ is nonnegative on the variety in equation (8). We fix a degree bound of $d = 1$ and let $u(X) = (1, X_{11}, X_{12}, X_{21}, X_{22})$. The dimensions of the matrices in the SDP (9) are $Q \in \mathbb{R}^{5 \times 5}$, $C \in \mathbb{R}^{4 \times 5}$. Solving the SDP leads to the following SOS certificate:

$$\begin{aligned} p(X) &= F(X) - g_1 h_1 - g_2 h_2 + g_3 h_3 - g_4 h_4 \\ F(X) &= (X_{21} - X_{12} + 1)^2 + (X_{11} - X_{22})^2 \\ g_1 &= X_{12} + 1, \quad g_2 = X_{21} + 1, \quad g_3 = X_{11} + X_{22}, \quad g_4 = X_{21} + X_{12} \\ h_1 &= X_{11}^2 + X_{21}^2 - 1, \quad h_2 = X_{12}^2 + X_{22}^2 - 1, \quad h_3 = X_{11}X_{12} + X_{21}X_{22}, \quad h_4 = \det(X) - 1 \end{aligned}$$

Compared to the sampling SDP in (7), the SDP in (9) has an additional matrix C , and also matrix Q is larger because it does not take advantage of the algebraic dependencies on the variety. A common approach to eliminate this unnecessary complexity is by using Gröbner bases, as described next.

A.2. Quotient ring SDP formulation. Quotient ring formulations take advantage of the algebraic relations derived from the equations h defining the variety to obtain a simpler SDP. The standard approach to obtain a quotient ring formulation requires a Gröbner basis of the ideal $I := \langle h \rangle$ [17]. We briefly explain the procedure now.

Let $\mathcal{R}' = \mathbb{R}[x]/I$ be the quotient ring of the ideal I . We denote by $\phi : \mathbb{R}[x] \rightarrow \mathcal{R}'$ the natural morphism onto \mathcal{R}' . Note that the image under this map of any polynomial of the form $g_j h_j$ is zero. Therefore, when we view the SDP in (9) in the quotient ring we obtain the following:

$$(10) \quad \begin{aligned} & \text{find} \quad Q \in \mathbb{R}^{N \times N}, \quad Q \succeq 0 \\ & \text{subject to} \quad \phi(p(x)) = Q \bullet \phi(u(x)u(x)^T) \end{aligned}$$

This SDP is simpler than the one in (9), as we have eliminated the matrix C .

The above SDP formulation requires methods to represent and compute the quotient ring \mathcal{R}' and the morphism ϕ . Given a Gröbner basis gb of I , there is a simple way to achieve this. Concretely, any polynomial $f \in \mathbb{R}[x]$ can be written in *normal form*, denoted as $\phi_{gb}(f)$, with respect to gb . This normal form map ϕ_{gb} is effectively representing the quotient ring. In addition, there is a natural monomial basis $u(x)$ to use, given by the standard monomials with respect to gb . Further improvements can be made to reduce the complexity of the SDP [19].

Example A.2. Consider again the case from Example 3.2. The following is a Gröbner basis of the equations in (8):

$$gb = \{\underline{X_{11}} - X_{22}, \underline{X_{12}} + X_{21}, \underline{X_{21}^2} + X_{22}^2 - 1\}.$$

The underlined terms are the leading monomials of the equations. Any polynomial $f \in \mathbb{R}[x]$ can be reduced by using the equations in gb in such a way that none of its terms are multiples of these leading monomials. The reduced form obtained, denoted as $\phi_{gb}(f)$, is said to be in normal form. For instance, the polynomial $p(X) = 4X_{21} - 2X_{11}X_{22} - 2X_{12}X_{21} + 3$ can be reduced to obtain $\phi_{gb}(p(X)) = 4X_{21} - 4X_{22}^2 + 5$.

We fix a degree bound of $d = 1$ and we set the vector $u(X) = (1, X_{21}, X_{22})$, which are the linear monomials that are in normal form (i.e., the standard monomials). After computing $\phi_{gb}(u(X)u(X)^T)$, we obtain the following SDP:

$$\begin{aligned} & \text{find} && Q \in \mathbb{R}^{3 \times 3}, \quad Q \succeq 0 \\ & \text{subject to} && 4X_{21} - 4X_{22}^2 + 5 = Q \bullet \begin{bmatrix} 1 & X_{21} & X_{22} \\ X_{21} & 1 - X_{22}^2 & X_{21}X_{22} \\ X_{22} & X_{21}X_{22} & X_{22}^2 \end{bmatrix} \end{aligned}$$

Solving the SDP leads to the SOS certificate $\phi_{gb}(p(X)) = \phi_{gb}((2X_{21} + 1)^2)$. Observe that the matrix dimension in the SDP is the same as for Example 3.2, and the SOS certificate obtained is also the same.

Let's see now that our sampling approach can be seen as a quotient ring formulation. The difference is that we use the radical ideal $J = \sqrt{I}$ and the coordinate ring $\mathcal{R} = \mathbb{R}[x]/J$. Given a sample set $Z = \{z_1, \dots, z_S\} \subseteq \mathcal{V}$ we describe the coordinate ring by using the *evaluation map* $\phi_Z : \mathbb{R}[x] \rightarrow \mathbb{C}^S$ such that

$$f \mapsto (f(z_1), f(z_2), \dots, f(z_S)).$$

Let's see that the kernel of ϕ_Z is given by J when restricted to a suitable linear subspace, and thus ϕ_Z gives a description of the ring $\mathbb{R}[x]/J$. It is clear that J is contained in the kernel of ϕ_Z . Consider now a linear subspace $\mathcal{L} \subseteq \mathbb{R}[x]$, such that $(\mathcal{L}, \mathcal{R})$ is poised. Note that the poisedness condition says that if $f \in \mathcal{L}$ is in the kernel of ϕ_Z , then f vanishes on \mathcal{V} , i.e., $f \in J$. This shows that the kernel of $\phi_Z|_{\mathcal{L}}$ is given by J .

Gröbner bases are the traditional approach to obtain quotient ring formulations. Our sampling methodology provides an alternative, which conveniently works over the radical ideal, i.e., it ignores multiplicities. In addition, sampling is simpler than computing Gröbner bases in many interesting cases.

REFERENCES

- [1] Daniel J Bates, Jonathan D Hauenstein, Andrew J Sommese, and Charles W Wampler. Bertini: Software for numerical algebraic geometry. Available at www.nd.edu/~sommese/bertini, 2006.
- [2] Daniel J Bates, Jonathan D Hauenstein, Andrew J Sommese, and Charles W Wampler. *Numerically solving polynomial systems with Bertini*, volume 25. SIAM, 2013.
- [3] Steven J Benson and Yinyu Ye. DSDP5: Software for semidefinite programming. *Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, Tech. Rep. ANL/MCS-P1289-0905*, 2005.

- [4] G. Blekherman, P. Parrilo, and R. Thomas. *Semidefinite optimization and convex algebraic geometry*, volume 13. MOS-SIAM Series on Optimization, 2013.
- [5] Moody T Chu and Nickolay T Trendafilov. The orthogonally constrained regression revisited. *Journal of Computational and Graphical Statistics*, 10(4):746–771, 2001.
- [6] Ciro Ciliberto. Geometric aspects of polynomial interpolation in more variables and of Waring’s problem. In *European Congress of Mathematics*, volume 201 of *Progress in Mathematics*, pages 289–316. Springer, 2001.
- [7] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000.
- [8] Jean-Charles Faugère. Finding all the solutions of Cyclic 9 using Gröbner basis techniques. In *Computer Mathematics - Proceedings of the Fifth Asian Symposium (ASCM 2001)*, volume 9, pages 1–12. World Scientific, 2001.
- [9] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at <http://www.math.uiuc.edu/Macaulay2/>.
- [10] Jens Carsten Jantzen. Nilpotent orbits in representation theory. In *Lie theory*, pages 1–211. Springer, 2004.
- [11] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.
- [12] Joseph M Landsberg and Jerzy Weyman. On the ideals and singularities of secant varieties of Segre varieties. *Bulletin of the London Mathematical Society*, 2007.
- [13] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009.
- [14] Zhang Liu and Lieven Vandenberghe. Low-rank structure in semidefinite programs derived from the KYP lemma. In *46th IEEE Conference on Decision and Control*, pages 5652–5659, 2007.
- [15] Johan Löfberg and Pablo A Parrilo. From coefficients to samples: a new approach to SOS optimization. In *43rd IEEE Conference on Decision and Control*, volume 3, pages 3154–3159, 2004.
- [16] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.
- [17] Pablo A Parrilo. Exploiting structure in sum of squares programs. In *42nd IEEE Conference on Decision and Control*, volume 5, pages 4664–4669, 2003.
- [18] Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96(2):293–320, 2003.
- [19] Frank Permenter and Pablo A Parrilo. Selecting a monomial basis for sums of squares programming over a quotient ring. In *IEEE 51st Annual Conference on Decision and Control*, pages 1871–1876, 2012.
- [20] Tae Roh, Bogdan Dumitrescu, and Lieven Vandenberghe. Multidimensional FIR filter design via trigonometric sum-of-squares optimization. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):641–650, 2007.
- [21] Tae Roh and Lieven Vandenberghe. Discrete transforms, semidefinite programming, and sum-of-squares representations of nonnegative polynomials. *SIAM Journal on Optimization*, 16(4):939–964, 2006.
- [22] Tomas Sauer. Polynomial interpolation in several variables: lattices, differences, and ideals. *Studies in Computational Mathematics*, 12:191–230, 2006.
- [23] Andrew John Sommese and Charles Weldon Wampler. *The Numerical solution of systems of polynomials arising in engineering and science*, volume 99. World Scientific, 2005.
- [24] Thorsten Theobald and Timo De Wolff. Approximating amoebas and coamoebas by sums of squares. *Mathematics of Computation*, 84(291):455–473, 2015.
- [25] Reha H Tütüncü, Kim C Toh, and Michael J Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical programming*, 95(2):189–217, 2003.

- [26] Jan Verschelde. Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software (TOMS)*, 25(2):251–276, 1999.
- [27] Thomas Viklands. *Algorithms for the weighted orthogonal Procrustes problem and other least squares problems*. PhD thesis, Umea University, Sweden, 2006.
- [28] Huan Wang, Shuicheng Yan, Dong Xu, Xiaoou Tang, and Thomas Huang. Trace ratio vs. ratio trace for dimensionality reduction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
- [29] Lei Hong Zhang and Ren Cang Li. Maximization of the sum of the trace ratio on the Stiefel manifold, I: Theory. *Science China Mathematics*, 57:2495–2508, 2014.
- [30] Lei Hong Zhang and Ren Cang Li. Maximization of the sum of the trace ratio on the Stiefel manifold, II: Computation. *Science China Mathematics*, 57:1–18, 2014.

LABORATORY FOR INFORMATION AND DECISION SYSTEMS (LIDS), MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE MA 02139, USA

E-mail address: diegcif@mit.edu

LABORATORY FOR INFORMATION AND DECISION SYSTEMS (LIDS), MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE MA 02139, USA

E-mail address: parrilo@mit.edu